Continuation methods and disjoint equilibria

Nathalie M. M. Cousin-Rittemard, Isabelle Gruais

CONTINUATION METHODS AND DISJOINT EQUILIBRIA

N.M.M. COUSIN–RITTEMARD, I. GRUAIS

Continuation methods are efficient to trace branches of fixed point solutions in parameter space as long as these branches are connected. However, the computation of isolated branches of fixed points is a crucial issue and require ad-hoc techniques. We suggest a modification of the standard continuation methods to determine these isolated branches more systematically. The so-called residue continuation method is a global homotopy starting from an arbitrary disjoint initial guess. Explicit conditions ensuring the quadratic convergence of the underlying Newton-Raphson process are derived and illustrated through several examples.

2000 Classification numbers: 37M20, 65p30.

Key words: Disjoint solution; continuation methods; error estimate; global homotopy; residue.

1. INTRODUCTION

The models of nonlinear physical phenomena depend on parameters. For example, in applications in fluid mechanics, the flows are described by a set of partial differential non-linear equations, i.e., the Navier–Stokes equations. In many cases, the transitions in model behavior (or bifurcation) are found as values of a particular parameter are changed. The Dynamical systems theory studies the common features of transitions in these nonlinear systems. One of the important basic issues of the bifurcation theory is the determination of fixed points (or steady states) of the system under investigation.

The first step of the bifurcation analysis is the discretization of the governing equations leading to a system of algebraic-differential equations of the form:

$$\frac{\partial \mathbf{u}}{\partial t} = \mathbf{A}(\mathbf{u}, \mu) \tag{1}$$

where $\mathbf{u} \in \mathbb{R}^n$ is the vector of the unknown quantities at the gridpoints, $\mu \in \mathbb{R}^{\mathbf{p}}$ is the vector the parameters, $\mathbf{A} : \mathbb{R}^{\mathbf{n}} \times \mathbb{R}^{\mathbf{p}} \to \mathbb{R}^{\mathbf{n}}$ is a nonlinear operator.

The fixed point solutions are determined by $\mathbf{A}(\mathbf{u}, \mu) = \mathbf{0}$. The branches of steady states are usually computed versus a control parameter using the so-called continuation methods [12, 19]. These continuation methods are very efficient when the branches are connected in the parameter space as techniques exist to switch between the branches in the vicinity of the bifurcation points. There is a rich literature on the application of these techniques in the fields of fluid or structural dynamics and mathematical analysis of the methodology [21, 4, 8, 17, 14, 1, 19, 15] for example. We refer to [15] for an up-to-date state of art.

One of the basic continuation methods is the *natural* continuation method. A branch of steady states of (1) is computed through infinitesimal increments of a control parameter $\mu \in \{\mu_1, \mu_2, ..., \mu_p\}$. Given a previously determined solution $(\mathbf{u}_0, \mu_0)$, the solution $(\mathbf{u}, \mu_0 + \delta\mu)$ is carried out with the Newton-Raphson method as it follows.

$$\mathbf{D_u A}^k \, \delta\mathbf{u^k} = -\mathbf{A^k} \quad , \quad \mathbf{u^k} \leftarrow \mathbf{u^k} + \delta\mathbf{u^k} \tag{2}$$

where $\delta\mu$ is a small increment with respect to μ_0 and $\mathbf{D_u A}^k$ is the Jacobian matrix of $\mathbf{A}^k$ with respect to $\mathbf{u}$ at the kth estimate $(\mathbf{u}^k, \mu_0 + \delta\mu)$. At regular points, it holds that

$$\text{rank}\left(\mathbf{D_u A}^k\right) = n. \tag{3}$$

The system arising from the Newton linearization is carried out using a classic solver.

However, the Jacobian is singular at the saddle-node bifurcation points and the natural continuation method cannot go around these bifurcation points. In the neigborhood of such a bifurcation point, one use the pseudo–arclength continuation [12]. In this method, an arclength parametrization of the solution branch is introduced of the form $(\mathbf{u}(s), \mu(s))$, where s is the arclength parameter. Because an additional degree of freedom is introduced, an additionnal equation is needed as the scalar normalization of the tangent along the branch :

$$\mathcal{N}\left(\mathbf{u}(s), \mu(s)\right) \equiv \mathbf{D_u N}.\delta\mathbf{u} + \mathbf{D}_\mu\mathbf{N} \, \delta\mu - \delta\mathbf{s} = \mathbf{0} \tag{4}$$

where $\mathbf{D}_\mu\mathbf{N}$ and $\mathbf{D_u N}$ are derivatives of the operator $\mathcal{N}$ with respect to μ and $\mathbf{u}$, respectively.

Given a solution $(\mathbf{u}(s_0), \mu(s_0))$, the solution $(\mathbf{u}(s_0 + \delta s), \mu(s_0 + \delta s))$ is determined again with the Newton-Raphson method through

$$\text{(5)} \qquad \begin{pmatrix} \mathbf{D_u A}^k & \mathbf{D}_\mu \mathbf{A}^k \\ \mathbf{D_u N}^{k^T} & \mathbf{D}_\mu \mathbf{N}^k \end{pmatrix} \begin{pmatrix} \delta \mathbf{u^k} \\ \delta \mu^k \end{pmatrix} = - \begin{pmatrix} \mathbf{A}^k \\ \mathcal{N}^k \end{pmatrix} \quad , \qquad \begin{pmatrix} \mathbf{u}^k \\ \mu^{\mathbf{k}} \end{pmatrix} \leftarrow \begin{pmatrix} \mathbf{u}^k + \delta \mathbf{u^k} \\ \mu^{\mathbf{k}} + \delta \mu^k \end{pmatrix}$$

where $\mathbf{D}_\mu \mathbf{A}^k$, $\mathbf{D}_\mu \mathbf{N}^k$ and $\mathbf{D_u N}^k$ are the derivatives of the operators at the current estimate $\left(\mathbf{u}^k, \mu^k\right)$. In practice, both natural continuation and pseudo-arclength continuation are implemented and a switch between the two schemes is based on a value of the slope $\mathbf{D}_\mu \mathbf{N}$ [5, 6, 7, 10]. Usually, one starts the continuation from a known trivial (or analytical) solution $(\mathbf{u}_0, \mu_0)$. However, in many applications, there exist branches of steady state solutions disconnected from the branch containing a trivial starting solution. These branches are the so-called *isolated* branches. Bifurcation theory in many cases may a priori indicate that there are isolated (or disjoint) branches of solutions. A typical example is the case of an imperfect pitchfork bifurcation, for example, occurring in the wind-driven ocean flows [3, 20].

There are basically three methods to compute these isolated branches. But it is not guaranteed that one will find all branches with either of these methods. Two of those are more or less trial and error while a more systematic approach is followed in the latter method.

(i) Transient integration.
In this approach, a set of initial conditions is chosen and a transient computation is started. If one is lucky, one of the initial conditions is in the attraction basin of a steady state on the isolated branch. Once found, one can continue tracing this branch using the continuation methods.

(ii) Isolated Newton-Raphson search.
One can also start a Newton-Raphson process uncoupled from the pseudo-arclength continuation from several chosen starting points. Since the convergence of the Newton-Raphson process is only quadratic in the vicinity of the steady state, this method may not be very efficient. But again, if one is lucky an isolated branch might be found.

(iii) Two-parameter continuation.
In several cases, a second parameter can be varied such that the targeted branch connects to an already known branch. An important example is the dynamical system with a particular symmetry and characterised by a pitchfork bifurcation for a specific value of the second parameter. Once the connection is

present, the isolated branch can be computed by restoring the second parameter to its original value.

As isolated branches are very important to determine, there is a need for more systematic methodology to find them. In this contribution, we propose a modification of the continuation methods which enables the determination of the isolated branches in §2. Explicit conditions ensuring the quadratic convergence of the Newton-Raphson process in the case of natural and pseudo-arclength continuation are derived in §3. The capabilities and efficiency of the method are illustrated in section §4, with help of several examples.

2. The Residue Continuation Method

The Residue Continuation Method is based on the global homotopy idea as pioneered by Keller in [13]. Let $\mathbf{u}^*$ be an initial guess to a fixed point of (1), then he considered the modified set of equations,

$$\mathbf{A}(\mathbf{u}) - e^{-\alpha(\mathbf{u})t}\ \mathbf{A}(\mathbf{u}^*) = 0$$

where $\alpha(\mathbf{u}) > 0$ and $0 < t < \infty$. When $t \to \infty$, the solution $\mathbf{u}(t)$, if it exists, approaches a fixed point of (1). Keller [13] showed that the choice of $\alpha(\mathbf{u})$ is crucial and he clarified the sharp conditions that this operator must satisfy on a tubular neighborhood of the path for the existence of $\mathbf{u}(t)$. As the proof is constructive, the choice of a particular $\alpha(\mathbf{u})$ for a given operator it clearly indicated.

In the present paper, a systematic approach is suggested. Let $(\mathbf{u}^*, \mu^*)$ be an initial guess of an isolated steady state $(\mathbf{u}_0, \mu_0)$ of (1), the idea of the residue continuation method is to solve the global homotopy :

$$\begin{aligned} \mathcal{H}(\mathbf{u}, \mu, \alpha) &\equiv \mathbf{A}(\mathbf{u}, \mu) - \alpha\ \mathbf{r} = 0\,, \\ \mathcal{K}(\mathbf{u}, \mu) &= k_{\mathbf{u}}.\mathbf{u} + k_\mu \mu = 0 \end{aligned} \tag{6}$$

where $\mathbf{r} = \mathbf{A}(\mathbf{u}^*, \mu^*)$ is the *residue*, α is the *residue parameter*. with $\mathcal{H} : \mathbb{R}^n \times \mathbb{R} \times [0,1] \to \mathbb{R}^n$ and $\mathcal{K} : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$. For a given residue $\mathbf{r}$, and assuming that $k_\mu \neq 0$, it follows from the Implicit Function Theorem that (6) can be written as

$$\begin{aligned} \mathcal{H}(\mathbf{u}(\alpha)) &\equiv \mathbf{A}(\mathbf{u}(\alpha), \mu(\mathbf{u}(\alpha))) - \alpha\ \mathbf{r} = 0 \\ \mathcal{K}(\mathbf{u}(\alpha)) &\equiv k_{\mathbf{u}}.\mathbf{u}(\alpha) + k_\mu\, \mu(\mathbf{u}(\alpha)) = 0 \end{aligned} \tag{7}$$

Let $(\alpha_\nu)_\nu$ be a real sequence such that $\alpha_\nu \in I \equiv [a,b] \subset \mathbb{R}$. For a given solution of the homotopy (6) $(\mathbf{u}_{\nu-1}, \mu_{\nu-1})$, we define the corresponding residue such that:

$$\mathbf{r}_{\nu-1} = \mathbf{A}(\mathbf{u}_{\nu-1}, \mu_{\nu-1}). \tag{8}$$

The Newton-Raphson method is used to solve the system of equations. For the natural continuation, the scheme can be written as

$$\mathbf{D_u}\mathcal{H}_\nu^k \, \delta\mathbf{u}_\nu^\mathbf{k} = -\mathcal{H}_\nu^\mathbf{k} \;\; , \quad \mathbf{u}_\nu^\mathbf{k} \leftarrow \mathbf{u}_\nu^\mathbf{k} + \delta\mathbf{u}_\nu^\mathbf{k} \tag{9}$$

while for the pseudo-arclength continuation, it becomes

$$\begin{pmatrix} \mathbf{D_u}\mathcal{H}_\nu^k & D_\alpha\mathcal{H}_\nu^k \\ \mathbf{D_u}\mathbf{N}_\nu^{kT} & D_\alpha\mathbf{N}_\nu^k \end{pmatrix} \begin{pmatrix} \delta\mathbf{u}_\nu^k \\ \delta\alpha_\nu^k \end{pmatrix} = - \begin{pmatrix} \mathcal{H}_\nu^k \\ \mathcal{N}_\nu^k \end{pmatrix} \;\; , \tag{10}$$

$$\begin{pmatrix} \mathbf{u}_\nu^k \\ \alpha_\nu^k \end{pmatrix} \leftarrow \begin{pmatrix} \mathbf{u}_\nu^k + \delta\mathbf{u}_\nu^k \\ \alpha_\nu^k + \delta\alpha_\nu^k \end{pmatrix}.$$

Using the pseudo-arclength continuation, the natural parameter μ_ν is such that $k_\mathbf{u}.\mathbf{u}_\nu + k_\mu\mu_\nu = 0$ and $\mathcal{H}_\nu^k \equiv \mathbf{A}_\nu^k - \alpha_\nu\mathbf{r}_{\nu-1}$. Furthermore, $\mathbf{D_u}\mathcal{H}_\nu^k$ and $D_\alpha\mathcal{H}_\nu^k$ are the Jacobian with respect to $\mathbf{u}$ and the residue parameter respectively, at the current estimate $(\mathbf{u}_\nu^k, \alpha_\nu^k)$ of the solution $(\mathbf{u}_\nu, \alpha_\nu)$ of (6).

Considering the equations (7) and (8), it follows that the residue parameter α_ν is such that:

$$|\alpha_\nu| = \frac{\|\mathbf{r}_\nu\|}{\|\mathbf{r}_{\nu-1}\|}. \tag{11}$$

Hence, it may be seen as the control parameter of the norm of the residue. The residue increases (respectively decreases) as long as $|\alpha| > 1$ (respectively $|\alpha| < 1$) and $\alpha = 1$ is a critical value corresponding to an extremum of the norm of the residue.

3. Convergence and estimates

In this section, a priori estimations of the convergence radius for the Newton–Raphson method are derived for the residue continuation scheme for both natural continuation (section § 3.1) and pseudo-arclength continuation (section § 3.2).

3.1. **Natural continuation.**

For any subdivision $(\alpha_\nu)_\nu$ of $I \subset \mathbb{R}$, $\mathbf{u}_\nu$ denotes the solution $\mathbf{u}(\alpha_\nu)$. Given an initial guess of the solution of homotopy (6) ($\mathbf{u}^0 \equiv \mathbf{u}_{\nu-1}$, $\mu_\nu^0 \equiv \mu_{\nu-1}$) with $\alpha \equiv \alpha_{\nu-1}$ and $\mathbf{r} \equiv \mathbf{r}_{\nu-2}$, the Newton-Raphson's scheme is written in the following equivalent form.

$$
\begin{array}{l}
\text{For } \nu = 1, \cdots, N \ , \\
\text{For } k = 0, \cdots, p_\nu - 1 \ , \\
\mathbf{D_u A}_\nu^k(\mathbf{u}_\nu^{k+1} - \mathbf{u}_\nu^k) = -\mathbf{A}_\nu^k + \alpha_\nu \mathbf{r}_{\nu-1} \quad , \quad \mathbf{u}_\nu^k \leftarrow \mathbf{u}_\nu^{k+1}
\end{array} \tag{12}
$$

Assuming that $\mathbf{D_u A}$ is nonsingular for every $\mathbf{u} \in \mathbb{R}^n$, the operators $\mathbf{A} : \mathbb{R}^n \mapsto \mathbb{R}^n$ and $\mathcal{H} : \mathbb{R}^n \mapsto \mathbb{R}^n$ are homeomorphisms. Therefore, for any α in some compact range $I_\nu \subset I \subset \mathbb{R}$, with ends α_ν and $\alpha_{\nu-1}$, the homotopy (6) admits a unique solution $\mathbf{u}(\alpha)$ such that:

$$
\begin{aligned}
\mathbf{u}(\alpha) &= \mathbf{A}^{-1}\left(\phi_\nu(\alpha)\mathbf{r}_{\nu-1}\right), \\
\phi_\nu(\alpha) &\equiv \frac{(1-\alpha_\nu)\alpha - (1-\alpha_{\nu-1})\alpha_\nu}{\alpha_{\nu-1} - \alpha_\nu}\, \mathbf{r}_{\nu-1}.
\end{aligned} \tag{13}
$$

As $\mathbf{A}^{-1}$ is continously differentiable, $\alpha \mapsto \mathbf{u}(\alpha)$ is continuous and piecewise C^1 while $\alpha \mapsto \mathbf{D_u A}(\mathbf{u}(\alpha))^{-1}$ is continuous on each I_ν. Therefore, assuming that the sequence $(\|\mathbf{r}_\nu\|)_\nu$ is bounded, there exists a constant $c > 0$ such that

$$
\|\mathbf{D_u A}(\mathbf{u}(\alpha))^{-1}\| \leq c, \quad \forall \alpha \in I_\nu \subset I \subset \mathbb{R} \tag{14}
$$

and it follows that

$$
\mathbf{u}'(\alpha) = \mathbf{D_u A}(\mathbf{u}(\alpha))^{-1} \left(\frac{(1-\alpha_\nu)}{\alpha_{\nu-1} - \alpha_\nu} \mathbf{r}_{\nu-1} \right), \quad \alpha \in I_\nu. \tag{15}
$$

We denote by $\mathcal{C}$ the limit curve of the Newton-Raphson process, defined as

$$
\mathcal{C} \equiv \{\mathbf{u}(\alpha) \in \mathbb{R}^n, \quad \alpha \in I \subset \mathbb{R}\}.
$$

As I is a compact convex and $\mathbf{u}$ is piecewise C^1 and continuous, there exists a compact and convex set $D \subset \mathbb{R}^n$ such that $\mathcal{C} \subset D$.

PROPOSITION 3.1. *For some constant $c' > 0$ we get:*

$$
\forall \mathbf{u} \in D : \quad \|\mathbf{D_u A}(\mathbf{u})^{-1}\| \leq c'. \tag{16}
$$

Proof. As $\mathbf{D_{uu} A}$ is continuous, it is also bounded on D and we define the constant $\kappa > 0$ as it follows:

$$
\|\mathbf{D_{uu} A}\| \leq \kappa, \quad \forall \mathbf{u} \in D. \tag{17}
$$

Because $\mathbf{D_u A}$ is continuous on D, it is also uniformly continuous on D and thus:

$$
\begin{array}{l}
\forall \varepsilon > 0, \ \exists \eta > 0, \ \forall \mathbf{u}, \mathbf{u}' \in D, \\
\|\mathbf{u} - \mathbf{u}'\| < \eta \Longrightarrow \|\mathbf{D_u A}(\mathbf{u}) - \mathbf{D_u A}(\mathbf{u}')\| \leq \varepsilon.
\end{array} \tag{18}
$$

Furthermore, given any $\mathbf{u}, \mathbf{u}' \in D$ and setting

$$
\mathbf{f}(\mathbf{t}) \equiv \mathbf{D_u A}\left(\mathbf{u} + \mathbf{t}(\mathbf{u}' - \mathbf{u})\right)
$$

we get, as f is continuously differentiable:

$$\mathbf{D_u A(u)} - \mathbf{D_u A(u')} = \mathbf{f(1)} - \mathbf{f(0)} = \int_{\mathbf{0}}^{\mathbf{1}} \mathbf{f'(t)dt}$$
$$= \int_0^1 \mathbf{D_{uu}A}(\mathbf{u} + t(\mathbf{u'} - \mathbf{u}))dt\ (\mathbf{u'} - \mathbf{u}).$$

Then, there holds

$$\forall \mathbf{u}, \mathbf{u'} \in D, \quad \|\mathbf{D_u A(u)} - \mathbf{D_u A(u')}\| =$$
$$\left\| \int_0^1 \mathbf{D_{uu}A}(\mathbf{u} + t(\mathbf{u'} - \mathbf{u}))dt\ (\mathbf{u'} - \mathbf{u}) \right\| \leq \kappa \|\mathbf{u} - \mathbf{u'}\|$$

so that we may choose $\eta = \dfrac{\varepsilon}{\kappa}$.

In addition, for every $\mathbf{u}^0 \in D$, we find

$$\|\mathbf{D_u A(u)}^{-1}\| \leq \frac{\|\mathbf{D_u A}(\mathbf{u}^0)^{-1}\|}{1 - \|\mathbf{D_u A}(\mathbf{u}^0)^{-1}\|\ \|\mathbf{D_u A(u)} - \mathbf{D_u A}(\mathbf{u}^0)\|}$$
$$\leq \frac{\|\mathbf{D_u A}(\mathbf{u}^0)^{-1}\|}{1 - \varepsilon \|\mathbf{D_u A}(\mathbf{u}^0)^{-1}\|}.$$

As a consequence, the uniform bound (16) holds with $c' = \dfrac{c}{1 - \varepsilon c}$ as soon as we choose

$$\varepsilon \in [0, \frac{1}{2c}].$$

This ends the proof. ■

Let $\beta_{0,\nu}$, $\eta_{0,\nu}$, $\gamma_{0,\nu}$, $t^{\pm}_{0,\nu}$ be defined as:

$$\|\mathbf{D_u A}(\mathbf{u}^0_\nu)^{-1}\| \leq \beta_{0,\nu} \equiv \frac{c}{1 - \varepsilon c} < +\infty, \tag{19}$$

$$\frac{|1 - \alpha_\nu|}{|\alpha_\nu|} \|(\mathbf{D_u A}(\mathbf{u}^0_\nu))^{-1} \mathbf{r}_\nu\| \leq \eta_{0,\nu} < +\infty, \tag{20}$$

$$\gamma_{0,\nu} \equiv \eta_{0,\nu}\ \beta_{0,\nu}\ \kappa, \tag{21}$$
$$t^{\pm}_{0,\nu} = \frac{1}{\kappa\ \beta_{0,\nu}} \left(1 \pm \sqrt{1 - 2\gamma_{0,\nu}}\right).$$

Following [11] (see [16]) and for each sequence $(\mathbf{u}^k_\nu)_k$, we introduce the sequences $\beta_{k,\nu}$, $\eta_{k,\nu}$, $\gamma_{k,\nu}$ and $t^{\pm}_{k,\nu}$ according to the following recurrence.

$$\gamma_{k,\nu} = \beta_{k,\nu}\ \eta_{k,\nu}\ \kappa,$$

$$\beta_{k+1,\nu} = \frac{\beta_{k,\nu}}{1 - \gamma_{k,\nu}},$$

$$\eta_{k+1,\nu} = \frac{\gamma_{k,\nu}\ \eta_{k,\nu}}{2(1-\gamma_{k,\nu})},$$

$$t^{\pm}_{k,\nu} = \frac{1}{\kappa\ \beta_{k,\nu}}\left(1 \pm \sqrt{1-2\gamma_{k,\nu}}\right).$$

Taking the limit $k \to \infty$ for the residue Newton-Raphson scheme (12), we have

$$\|\mathbf{u}^0_\nu - \mathbf{u}(\alpha_\nu)\| \equiv \|\mathbf{u}^0_\nu - \mathbf{u}_\nu\| \leq t^-_{0,\nu} = t^-_{p_{\nu-1},\nu-1} \tag{22}$$

and, according to the definition of $\mathbf{u}^0_\nu \equiv \mathbf{u}_{\nu-1}$:

$$\|\mathbf{u}^0_\nu - \mathbf{u}(\alpha_{\nu-1})\| \equiv \|\mathbf{u}^{p_{\nu-1}}_{\nu-1} - \mathbf{u}_{\nu-1}\| \leq 2\eta_{p_{\nu-1},\nu-1} \leq \frac{2\eta_{0,\nu-1}}{2^{p_{\nu-1}}}. \tag{23}$$

Kantorovich's Theorem then reads:

COROLLARY 3.2. *For each index ν, the sequence $(\mathbf{u}^k_\nu)_k$ generated by the scheme (12) converges to the unique solution $\mathbf{u}_\nu \equiv \mathbf{u}(\alpha_\nu)$ of the system:*

$$\mathbf{A}(\mathbf{u}_\nu) - \alpha_\nu \mathbf{r}_{\nu-1} = 0 \quad , \quad \mathbf{r}_{\nu-1} \equiv \mathbf{A}(\mathbf{u}_{\nu-1})$$

in the open ball $B(\mathbf{u}^0_\nu, t^+_{0,\nu})$ with

$$t^+_{0,\nu} = \frac{1}{\kappa\beta_{0,\nu}}\left(1 + \sqrt{1-2\gamma_{0,\nu}}\right)$$

and where κ, $\beta_{0,\nu}$ and $\gamma_{0,\nu}$ are defined as in (17), (19) and (21) respectively.

A sufficient condition for the convergence of (12) can now be stated as it follows.

PROPOSITION 3.3. *A sufficient condition for the sequence $(\mathbf{u}^k_\nu)_{k\geq 0}$ to converge towards $\mathbf{u}(\alpha_\nu)$ is that α_ν satisfies*

$$0 < \frac{|1-\alpha_\nu|}{|\alpha_\nu|}\|\mathbf{r}_\nu\| < \frac{1}{2c}\min\left(\frac{3-\sqrt{5}}{2\kappa c}, t^-_{0,\nu}\right) \equiv \Lambda_\nu,$$

where the constant c has been defined in (14).

Proof. We are seeking for a condition compatible with (22) and ensuring (18) with $\eta = \dfrac{\varepsilon}{\kappa}$, $\mathbf{u} = \mathbf{u}^0_\nu$ and $\mathbf{u}' = \mathbf{u}(\alpha_\nu)$ that is

$$||\mathbf{u}^0_\nu - \mathbf{u}(\alpha_\nu)|| < \frac{\varepsilon}{\kappa}. \tag{24}$$

We first notice that the following inequality remains true:

$$\|\mathbf{u}^0_\nu - \mathbf{u}(\alpha_\nu)\| \leq \|\mathbf{u}^0_\nu - \mathbf{u}(\alpha_{\nu-1})\| + \|\mathbf{u}(\alpha_{\nu-1}) - \mathbf{u}(\alpha_\nu)\|. \tag{25}$$

The first term of the right hand side is estimated according to (23) as it follows:

$$\frac{2\eta_{0,\nu-1}}{2^{p_{\nu-1}}} < \frac{\varepsilon}{2\kappa},$$

which holds true if one chooses $p_{\nu-1}$ such that

$$2^{p_{\nu-1}} > \frac{4\eta_{0,\nu-1}\kappa}{\varepsilon}.$$

An estimation of the second term of the right hand side of (25) is needed, says as:

$$\|\mathbf{u}(\alpha_{\nu-1}) - \mathbf{u}(\alpha_\nu)\| = \| \int_{\alpha_{\nu-1}}^{\alpha_\nu} \mathbf{u}'(\alpha) d\alpha \|.$$

Recalling that

$$\forall \alpha \in I_\nu : \quad \mathbf{A}(\mathbf{u}(\alpha)) = \frac{(1-\alpha_\nu)\alpha - (1-\alpha_{\nu-1})\alpha_\nu}{\alpha_{\nu-1} - \alpha_\nu} \mathbf{r}_{\nu-1} \tag{26}$$

where

$$I_\nu \equiv [\min(\alpha_{\nu-1}, \alpha_\nu), \max(\alpha_{\nu-1}, \alpha_\nu)], \tag{27}$$

and taking into account (15) yields to the following estimation.

$$\|\mathbf{u}(\alpha_{\nu-1}) - \mathbf{u}(\alpha_\nu)\| = \frac{|1-\alpha_\nu|}{|\alpha_{\nu-1} - \alpha_\nu|} \left\| \int_{\alpha_{\nu-1}}^{\alpha_\nu} \mathbf{D}_{\mathbf{u}} \mathbf{A}(\mathbf{u}(\alpha))^{-1} \mathbf{r}_{\nu-1} d\alpha \right\| \leq$$

$$\frac{|1-\alpha_\nu|}{|\alpha_{\nu-1} - \alpha_\nu|} \left| \int_{\alpha_\nu}^{\alpha_{\nu-1}} \left\| \mathbf{D}_{\mathbf{u}} \mathbf{A}(\mathbf{u}(\alpha))^{-1} \mathbf{r}_{\nu-1} \right\| d\alpha \right| \leq$$

$$c|1-\alpha_\nu| \|\mathbf{r}_{\nu-1}\| = c\frac{|1-\alpha_\nu|}{|\alpha_\nu|} \|\mathbf{r}_\nu\|.$$

One may choose α_ν such that

$$c\frac{|1-\alpha_\nu|}{|\alpha_\nu|} \|\mathbf{r}_\nu\| = \frac{2\eta_{0,\nu-1}}{2^{p_{\nu-1}}} < \frac{\varepsilon}{2\kappa}. \tag{28}$$

After substitution into (25), we get (24), says as

$$\|\mathbf{u}_\nu^0 - \mathbf{u}(\alpha_\nu)\| \leq 2c\frac{|1-\alpha_\nu|}{|\alpha_\nu|} \|\mathbf{r}_\nu\| < \frac{\varepsilon}{\kappa}.$$

As for (22), the same argument with $t_{0,\nu}^-$ instead of $\frac{\varepsilon}{\kappa}$ shows that :

$$||\mathbf{u}_\nu^0 - \mathbf{u}(\alpha_\nu)|| \leq 2c\frac{|1-\alpha_\nu|}{|\alpha_\nu|} ||\mathbf{r}_\nu|| < t_{0,\nu}^- \tag{29}$$

in accordance with (22). Moreover, a sufficient condition for the convergence to hold is such that:

$$0 < \eta_{0,\nu} < \frac{1 - \varepsilon c}{2\kappa c}. \tag{30}$$

Indeed, from [16], a sufficient condition for Newton's method to converge at the given step ν is that $\gamma_{0,\nu} < \frac{1}{2}$. Arguing as in the proof of (16), with $\mathbf{u}$ and $\mathbf{u}'$ replaced by $\mathbf{u}^0_\nu$ and $\mathbf{u}(\alpha_\nu)$ respectively, and taking advantage of the estimate (24), we get:

$$\|\mathbf{D_u A}(\mathbf{u}^0_\nu)^{-1}\| \leq \frac{\|\mathbf{D_u A}(\mathbf{u}(\alpha_\nu))^{-1}\|}{1 - \|\mathbf{D_u A}(\mathbf{u}^0_\nu) - \mathbf{D_u A}(\mathbf{u}(\alpha_\nu))\|\|\mathbf{D_u A}(\mathbf{u}(\alpha_\nu))^{-1}\|}$$

$$\leq \frac{c}{1 - \varepsilon c} = c' \equiv \beta_{0,\nu}. \tag{31}$$

Then, due to the definition (21), the condition $\gamma_{0,\nu} < \frac{1}{2}$ becomes

$$0 < \eta_{0,\nu} < \frac{1}{2\beta_{0,\nu}\kappa} = \frac{1 - \varepsilon c}{2c\kappa}, \tag{32}$$

which is (30). Notice that (20) holds with

$$\|(\mathbf{D_u A}(\mathbf{u}^0_\nu))^{-1}\mathbf{r}_\nu\| \leq c'\|\mathbf{r}_\nu\|.$$

Taking into account the equation (31), a sufficient condition for (20) to hold reads

$$\frac{|1 - \alpha_\nu|}{|\alpha_\nu|}\|\mathbf{r}_\nu\| < \frac{\eta_{0,\nu}}{c'}$$

which, after substitution of (30)-(31), leads to:

$$\frac{|1 - \alpha_\nu|}{|\alpha_\nu|}\|\mathbf{r}_\nu\| < \frac{1}{2\kappa}\left(\frac{1 - \varepsilon c}{c}\right)^2.$$

Comparing with (28) and (29), we get:

$$0 < \frac{|1 - \alpha_\nu|}{|\alpha_\nu|}\|\mathbf{r}_\nu\| < \frac{1}{2}\min\left(\frac{\varepsilon}{\kappa c}, \frac{t^-_{0,\nu}}{c}, \frac{1}{\kappa}\left(\frac{1 - \varepsilon c}{c}\right)^2\right).$$

Direct computation yields:

$$\frac{\varepsilon}{\kappa c} < \frac{1}{\kappa}\left(\frac{1 - \varepsilon c}{c}\right)^2 \iff \varepsilon^2 - \frac{3\varepsilon}{c} + \frac{1}{c^2} > 0.$$

This is realised as soon as:

$$0 < \varepsilon < \frac{3 - \sqrt{5}}{2c},$$

thus finishing the proof. ■

3.2. **Pseudo-arclength continuation.** The pseudo-arclength continuation could be used in cases of a non-regular Jacobian $\mathbf{D_u A}$.

For any given $\mathbf{r}$ and assuming that $\mathrm{rank}(\mathbf{D_u A}) = n-1$, we introduce the operator $\mathcal{F} : \mathbb{R}^{n+1} \mapsto \mathbb{R}^n$, as

$$\mathcal{F}(\mathbf{u}(s), \alpha(s); s) \equiv \begin{pmatrix} \mathcal{H}(\mathbf{u}(s), \alpha(s)) \\ \mathcal{N}(\mathbf{u}(s), \alpha(s); s) \end{pmatrix}.$$

Following the Implicit Function Theorem and for any given $\mathbf{r}$, the global homotopy (6) can be written as

$$\begin{aligned} \mathcal{F}(\mathbf{u}(s), \alpha(s); \mathbf{r}) &= 0, \\ k_{\mathbf{u}} \cdot \mathbf{u}(s) + k_\mu \mu(\mathbf{u}(s)) &= 0. \end{aligned} \tag{33}$$

For some fixed $s_\nu > 0$, consider Newton's scheme (10) written in the equivalent form:

For $\nu = 1, \cdots, N$,
For $k = 0, \cdots, p_\nu - 1$,
$\mathbf{D_u A}_\nu^k(\mathbf{u}_\nu^{k+1} - \mathbf{u}_\nu^k) - (\alpha_\nu^{k+1} - \alpha_\nu^k)\mathbf{r}_{\nu-1} = -\mathbf{A}_\nu^k + \alpha_\nu^k \mathbf{r}_{\nu-1}$,
$\mathbf{D_u N}_\nu^k(\mathbf{u}_\nu^{k+1} - \mathbf{u}_\nu^k) + \mathbf{D}_\alpha \mathbf{N}_\nu^k(\alpha_\nu^{k+1} - \alpha_\nu^k) = -\mathcal{N}_\nu^k$,
$(\mathbf{u}_\nu^k, \alpha_\nu^k) \leftarrow (\mathbf{u}_\nu^{k+1}, \alpha_\nu^{k+1})$.

For any $1 \le \nu \le N$, the corresponding value of the control parameter μ_ν is such that

$$k_{\mathbf{u}} \,.\, (\mathbf{u}_\nu - \mathbf{u}_{\nu-1}) + k_\mu(\mu_\nu - \mu_{\nu-1}) = 0$$

where the initialization point $(\mathbf{u}_\nu^0, \alpha_\nu^0) \equiv (\mathbf{u}_{\nu-1}, \alpha_{\nu-1})$ is taken to be solution of (6) with $s \equiv s_{\nu-1}$ and $\mathbf{r} \equiv \mathbf{r}_{\nu-2}$.

In the sequel, we assume that the matrix

$$\mathcal{B}(\mathbf{u}, \alpha) \equiv \begin{pmatrix} \mathbf{D_u A}(\mathbf{u}) & -\mathbf{r} \\ \mathbf{D_u N}(\mathbf{u}, \alpha) & \mathbf{D}_\alpha \mathbf{N}(\mathbf{u}, \alpha) \end{pmatrix}$$

is nonsingular for every $(\mathbf{u}, \alpha) \in \mathbb{R}^n \times \mathbb{R}$ and we assume that there is a constant $c > 0$ such that

$$\|\mathcal{B}^{-1}(\mathbf{u}, \alpha)\| \le c, \quad \forall (\mathbf{u}, \alpha) \in \mathbb{R}^n \times \mathbb{R}.$$

Since by construction, $\mathcal{F}(\cdot, \cdot; s, \mathbf{r})$ is a homeomorphism $\mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n \times \mathbb{R}$, for every $s > 0$, (33) admits a unique solution denoted by $(\mathbf{u}(s), \alpha(s))$:

$$\begin{pmatrix} \mathbf{u}(s) \\ \alpha(s) \end{pmatrix} = \mathcal{F}(\cdot, \cdot; s, \mathbf{r})^{-1} \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

As $\mathcal{F}^{-1}$ is continuously differentiable, $s \mapsto \mathcal{F}(\cdot, \cdot; s, \mathbf{r})^{-1}$ is of class C^1 as well as $s \mapsto (\mathbf{u}(s), \alpha(s))$. In particular, there exists a constant $c > 0$ such that

$$\|\mathcal{B}(\mathbf{u}(s), \alpha(s))^{-1}\| \le c, \quad \forall s \in \mathbb{R}, \tag{34}$$

and there holds

$$\begin{pmatrix} \mathbf{u}'(s) \\ \alpha'(s) \end{pmatrix} = \mathcal{B}(\mathbf{u}(s), \alpha(s))^{-1} \begin{pmatrix} 0 \\ -\mathbf{D}_s \mathcal{N}(\mathbf{u}(\mathbf{s}), \alpha(s)) \end{pmatrix}. \tag{35}$$

Consider the sequence $(\mathbf{y}_\nu^k)_{\nu,k} \equiv (\mathbf{u}_\nu^k, \alpha_\nu^k)_{\nu,k}$ defined for the subdivision of $I \subset \mathbb{R}$:

$$s_0 < s_1 < \cdots < s_N, \quad h_\nu = s_\nu - s_{\nu-1} > 0,$$

by the following scheme.

$$\begin{aligned} (36)\quad \text{For } \nu &= 1, \cdots, N, \\ \mathbf{y}_\nu^0 &= \mathbf{y}_{\nu-1}, \\ \mathbf{y}_\nu^{k+1} &= \mathbf{y}_\nu^k - \mathcal{B}(\mathbf{y}_\nu^k)^{-1} \begin{pmatrix} \mathbf{A}_\nu^k - \alpha_\nu^k \mathbf{r}_{\nu-1} \\ \mathcal{N}_\nu^k \end{pmatrix};\ k = 0, \cdots, p_\nu - 1, \\ \mathbf{y}_\nu &= \mathbf{y}_\nu^{p_\nu}, \quad \mathbf{r}_\nu = \mathbf{A}(\mathbf{u}_\nu). \end{aligned}$$

Defining the set

$$\mathcal{C} \equiv \{(\mathbf{u}(s), \alpha(s)) \in \mathbb{R}^n \times \mathbb{R}, \quad s \in [s_0, s_N]\},$$

then there exists a compact convex set $D \subset \mathbb{R}^n \times \mathbb{R}$ such that $\mathcal{C} \subset D$. As $\mathbf{D_y}\mathcal{B}$ is continuous, it is also bounded on D and we have

$$\|\mathbf{D_y}\mathcal{B}(\mathbf{y})\| \leq \kappa, \quad \forall \mathbf{y} \equiv (\mathbf{u}, \alpha) \in D,$$

for some constant $\kappa > 0$. As $\mathcal{B}$ is continuous on D, it is also uniformly continuous on D, that is:

$$\begin{aligned} (37)\qquad \forall \varepsilon > 0, \quad \exists \eta > 0, \quad \forall \mathbf{y}, \mathbf{y}' \in D, \\ \|\mathbf{y} - \mathbf{y}'\| < \eta \Longrightarrow \|\mathcal{B}(\mathbf{y}) - \mathcal{B}(\mathbf{y}')\| \leq \varepsilon. \end{aligned}$$

Furthermore,

$$\|\mathcal{B}(\mathbf{y}) - \mathcal{B}(\mathbf{y}')\| =$$

$$\left\| \int_0^1 \mathbf{D_y}\mathcal{B}(\mathbf{y} + t(\mathbf{y}' - \mathbf{y}))dt\ (\mathbf{y}' - \mathbf{y}) \right\| \leq \kappa \|\mathbf{y} - \mathbf{y}'\|, \quad \forall \mathbf{y}, \mathbf{y}' \in D$$

so that we may choose $\eta = \dfrac{\varepsilon}{\kappa}$. Notice that, due to (37),

$$\|\mathcal{B}(\mathbf{y})^{-1}\| \leq \frac{\|\mathcal{B}(\mathbf{y}^0)^{-1}\|}{1 - \|\mathcal{B}(\mathbf{y}^0)^{-1}\| \|\mathcal{B}(\mathbf{y}) - \mathcal{B}(\mathbf{y}^0)\|} \leq$$

$$\frac{\|\mathcal{B}(\mathbf{y}^0)^{-1}\|}{1 - \varepsilon \|\mathcal{B}(\mathbf{y}^0)^{-1}\|} \leq \frac{c}{1 - \varepsilon c} \equiv c'$$

which makes sense as soon as we choose:

$$(38)\qquad \varepsilon \in [0, \frac{1}{2c}].$$

With each sequence $(\mathbf{y}_\nu^k)_k$, we may associate the quantities $\beta_{k,\nu}$, $\eta_{k,\nu}$, $\gamma_{k,\nu}$ and $t_{k,\nu}^-$ as:

$$\|\mathbf{D}_{\mathbf{y}}\mathcal{B}(\mathbf{y})\| \leq \kappa, \quad \forall \mathbf{y} \in D, \tag{39}$$

$$\|\mathcal{B}(\mathbf{y}_\nu^0)^{-1}\| \leq c' \equiv \beta_{0,\nu} < +\infty, \tag{40}$$

$$\left\| (\mathcal{B}(\mathbf{y}_\nu^0))^{-1} \left(\frac{(1-\alpha_\nu^0)}{\alpha_\nu^0} \mathbf{r}_\nu, 0 \right)^T \right\| \leq \eta_{0,\nu} < +\infty,$$

$$\gamma_{0,\nu} \equiv \eta_{0,\nu}\beta_{0,\nu}\kappa \leq \frac{1}{2}, \tag{41}$$

$$t_{k,\nu}^{\pm} = \frac{1}{\kappa\beta_{k,\nu}} \left(1 \pm \sqrt{1-2\gamma_{k,\nu}}\right), \tag{42}$$

$$\gamma_{k,\nu} = \beta_{k,\nu}\eta_{k,\nu}\kappa,$$

$$\beta_{k+1,\nu} = \frac{\beta_{k,\nu}}{1-\gamma_{k,\nu}},$$

$$\eta_{k+1,\nu} = \frac{\gamma_{k,\nu}\eta_{k,\nu}}{2(1-\gamma_{k,\nu})},$$

where we took into account that

$$\mathcal{N}_\nu^0 = \mathbf{D}_{\mathbf{u}}\mathbf{A}_\nu^0\delta\mathbf{u}_\nu^0 + \mathbf{D}_\mu\mathbf{A}_\nu^0\delta\mu_\nu^0$$

$$\equiv \mathbf{D}_{\mathbf{u}}\mathbf{A}_\nu^0(\mathbf{u}_\nu^0 - \mathbf{u}_{\nu-1}^{p_{\nu-1}}) + \mathbf{D}_\alpha\mathbf{A}_\nu^0(\alpha_\nu^0 - \alpha_{\nu-1}^{p_{\nu-1}}) = 0$$

and where

$$\mathbf{y}_\nu^{k+1} = \mathbf{y}_\nu^k - \mathcal{B}(\mathbf{y}_\nu^k)^{-1}\mathcal{F}(\mathbf{y}_\nu^k; s_\nu, \mathbf{r}_{\nu-1}).$$

Now, the same arguments as in the previous section with $\mathbf{D}_{\mathbf{u}}\mathbf{A}$, $\mathbf{u}$, α replaced by $\mathcal{B}$, $\mathbf{y}$, s respectively yield:

COROLLARY 3.4. *For each index ν, the sequence $(\mathbf{y}_\nu^k)_k$ generated by the scheme (36) converges to the unique solution $\mathbf{y}_\nu \equiv \mathbf{y}(s_\nu) = (\mathbf{u}(s_\nu), \alpha(s_\nu))$ of the system:*

$$\mathbf{A}(\mathbf{u}_\nu) - \alpha_\nu\mathbf{r}_{\nu-1} = 0, \quad \mathcal{K}(\mathbf{u}_\nu) = 0, \quad \mathcal{N}(\mathbf{y}_\nu; s_\nu) = 0, \quad \mathbf{r}_{\nu-1} \equiv \mathbf{A}(\mathbf{u}_{\nu-1})$$

in the open ball $B(\mathbf{y}_\nu^0, t_{0,\nu}^+)$ with

$$t_{0,\nu}^+ = \frac{1}{\kappa\beta_{0,\nu}} \left(1 + \sqrt{1-2\gamma_{0,\nu}}\right)$$

and where κ, $\beta_{0,\nu}$ and $\gamma_{0,\nu}$ are defined as in (39), (40) and (41) respectively.

PROPOSITION 3.5. *A sufficient condition for the sequence* $(\mathbf{y}_\nu^k)_{k\geq 0}$ *to converge towards* $\mathbf{y}(s_\nu)$ *is that*

$$0 < h_\nu < \frac{1}{2c\|\mathbf{D}_s\mathcal{N}\|}\min\left(\frac{1}{2\kappa}, t_{0,\nu}^-\right) \quad \textit{and} \tag{43}$$

$$0 < \frac{|1-\alpha_\nu^0|}{|\alpha_\nu^0|}\|\mathbf{r}_\nu\| < \frac{1}{2\kappa c^2} \equiv \Lambda_\nu, \tag{44}$$

where the constant c *has been defined in (34).*

Proof. The proof follows the same lines as that of proposition 3.3. The above theory applies if the condition (37) is realised with $\eta = \dfrac{\varepsilon}{\kappa}$, $\mathbf{y} = \mathbf{y}_\nu^0$, $\mathbf{y}\,' = \mathbf{y}(\alpha_\nu)$, that is:

$$\|\mathbf{y}_\nu^0 - \mathbf{y}(\alpha_\nu)\| < \frac{\varepsilon}{\kappa}.$$

As a consequence, the analogue of (18) will hold true. To achieve this, notice that

$$\|\mathbf{y}_\nu^0 - \mathbf{y}(s_\nu)\| \leq \|\mathbf{y}_\nu^0 - \mathbf{y}(s_{\nu-1})\| + \|\mathbf{y}(s_{\nu-1}) - \mathbf{y}(s_\nu)\|. \tag{45}$$

The first term in the right-hand side is handled by the same argument as for (23), namely:

$$\|\mathbf{y}_\nu^0 - \mathbf{y}(s_{\nu-1})\| \equiv \|\mathbf{y}_{\nu-1}^{p_{\nu-1}} - \mathbf{y}_{\nu-1}\| \leq 2\eta_{p_{\nu-1},\nu-1} \leq \frac{2\eta_{0,\nu-1}}{2^{p_{\nu-1}}}.$$

Therefore, we may choose $p_{\nu-1}$ so that

$$\frac{2\eta_{0,\nu-1}}{2^{p_{\nu-1}}} < \frac{\varepsilon}{2\kappa}$$

that is,

$$2^{p_{\nu-1}} > \frac{4\eta_{0,\nu-1}\kappa}{\varepsilon},$$

which fixes $p_{\nu-1}$ and also $\mathbf{y}_\nu^0 = \mathbf{y}_{\nu-1}^{p_{\nu-1}}$. The second term in the right-hand side of (45) remains to be estimated. Indeed, the equation (35) yields:

$$\|\mathbf{y}(s_{\nu-1}) - \mathbf{y}(s_\nu)\| = \left\|\int_{s_{\nu-1}}^{s_\nu} \mathbf{y}'(s)ds\right\| = \left\|\int_{s_{\nu-1}}^{s_\nu} \mathcal{B}(\mathbf{y}(s))^{-1}\left(0, -\mathbf{D}_s\mathcal{N}(\mathbf{y}(\mathbf{s}))\right)^T ds\right\|$$

$$\leq \int_{s_{\nu-1}}^{s_\nu} \left\|\mathcal{B}(\mathbf{y}(s))^{-1}\right\| \|\mathbf{D}_s\mathcal{N}\| ds \leq c\|\mathbf{D}_s\mathcal{N}\| h_\nu.$$

Considering the step ν, we may choose h_ν in such a way that

$$c\|\mathbf{D}_s\mathcal{N}\| h_\nu = \frac{2\eta_{0,\nu-1}}{2^{p_{\nu-1}}}$$

while, by construction of $p_{\nu-1}$, we get:

$$c\|\mathbf{D}_s\mathcal{N}\| h_\nu = \frac{2\eta_{0,\nu-1}}{2^{p_{\nu-1}}} < \frac{\varepsilon}{2\kappa} < \frac{1}{4\kappa c}, \tag{46}$$

the last inequality being a consequence of (38).
The analogue of (22) reads

$$\|\mathbf{y}_\nu^0 - \mathbf{y}(s_\nu)\| \equiv \|\mathbf{y}_\nu^0 - \mathbf{y}_\nu\| \leq t_{0,\nu}^- = t_{p_{\nu-1},\nu-1}^-.$$

Then, repeating the argument leading to (46) with $t_{0,\nu}^-$ instead of $\frac{\varepsilon}{\kappa}$ yields

$$0 < h_\nu < \frac{t_{0,\nu}^-}{2c\|\mathbf{D}_s\mathcal{N}\|}.$$

Combining with (46), we get:

$$0 < h_\nu < \frac{1}{2c\|\mathbf{D}_s\mathcal{N}\|}\min\left(\frac{1}{2\kappa}, t_{0,\nu}^-\right),$$

that is (43).
Moreover, arguing as in the previous section to establish (31)–(32), we find that the requirement (30) remains true, that is

$$0 < \eta_{0,\nu} < \frac{1-\varepsilon c}{2\kappa c}.$$

Recall that

$$\eta_{0,\nu} \geq \left\| (\mathcal{B}(\mathbf{y}_\nu^0))^{-1} \left(\frac{(1-\alpha_\nu^0)}{\alpha_\nu^0}\mathbf{r}_\nu, 0 \right)^T \right\|$$

as well as the definition (40) of $\beta_{0,\nu}$, to deduce that $\eta_{0,\nu}$ may be chosen as:

$$\left\| (\mathcal{B}(\mathbf{y}_\nu^0))^{-1} \left(\frac{(1-\alpha_\nu^0)}{\alpha_\nu^0}\mathbf{r}_\nu, 0 \right)^T \right\| \leq \frac{c}{(1-\varepsilon c)} \left(\frac{|1-\alpha_\nu^0|}{|\alpha_\nu^0|}\|\mathbf{r}_\nu\| \right) \equiv \eta_{0,\nu}.$$

We must have:

$$\frac{|1-\alpha_\nu^0|}{|\alpha_\nu^0|}\|\mathbf{r}_\nu\| = \eta_{0,\nu}\frac{(1-\varepsilon c)}{c} < \frac{(1-\varepsilon c)^2}{2\kappa c^2} < \frac{1}{2\kappa c^2}$$

which is (44). ■

4. Numerical experiments

In this section, we illustrate the residue continuation method along three examples. The first example addresses the problem of finding a fixed point solution of a scalar equation (i.e. $\mu \in \mathbb{R}$ and $u \in \mathbb{R}$) starting from an arbitrary initial guess. In the second and the third examples again a scalar equation is used. It is shown how the method is able to reach an isolated solution ans the convergence properties of the method. The last example is an application to a multidimensional set of equations encountered in the analysis of the stability of mechanical structures [22, 9]. These three examples illustrate the key ideas and convergence properties of the residue continuation method.

4.1. **Starting from a remote estimate.** This paragraph illustrates how to find a fixed point solution on a branch using the residue continuation method, starting from a disjoint guess. Consider the scalar equation

$$\mathbf{A}(u,\mu) = (u-1)^2 + \mu + 1 \tag{47}$$

with $\mathbf{A} : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$. The branch of steady states is carried out as it follows. We consider $(u^* = 55, \mu^* = -10)$ as the initial guess marked with a square in Figure 1(a). It is a far too coarse guess for the classical continuation methods. Nevertherless, using the residue continuation method with $k_{\mathbf{u}} = 0$, $k_\mu = 1$, a path is found (denoted by pt#1 in Figure 1(a)) from this remote estimate to the corresponding solution on the branch marked with a dot. Starting from this point, the branch of steady states can be determined using the classical pseudo-arclength continuation method (solid and dash-dotted lines). From the initial remote estimate and choosing other directions, other paths can be taken. An example where $k_{\mathbf{u}} = 0.1$, $k_\mu = 1$ is shown in Figure 1(b).

The example illustrates that for some given discretized operator, one can start from a remote initial guess and find systematically a fixed point solution of the operator, using the residue continuation method. One could have alternatively started the Newton-Raphson scheme from the initial estimate, but this is not guaranteed to converge. The method can thus be applied for problems where no trivial or analytical solution is known, and determine a first fixed point solution. In case of one does not want to compute solutions from the trivial state because of computational constraints, the method here can provide an efficient initial nontrivial solution.

4.2. **Location of an isolated branch.** Suppose a dynamical system has more than a single branch of fixed point solutions and that only one of these branches has been computed (for example, by starting from a trivial solution). In this example, we illustrate how the residue continuation method can be used to compute the other branches. Again, we take a simple scalar equation, in this case

$$\begin{aligned} \mathbf{A}(u,\mu) \;\equiv\; & \left((u-1)^2 + \mu + 1\right)\left((u-10)^2 - \mu - 5\right) \\ & \left((u-7)^2 + \mu + 10\right) \end{aligned} \tag{48}$$

where $\mathbf{A} : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$. As shown in Figure 2(a), there are three branches of steady states labelled #1, #2 and #3 and corresponding to the three factors of the right hand side of the system (48). The branches #1 and #3 are connected through a transcritical bifurcation while the branch #2 is an isolated branch. Our target is to compute

a point on the branch #2 starting from a point of the branch #1 as a remote estimate.

Using the procedure as in section 4.1, we first reach a point on branch #3 through the residue continuation path pt#1 (Figure 2(a)). From this fixed point, the branch #3 is computed using standard pseudo-arclength continuation rounding the saddle-node bifurcation at $u = 7$, $\mu = -10$. In a typical application, one would detect the transcritical bifurcation at $(u = 4.75, \mu^* = -15)$ and then use a branch switching method to calculate the branch #1. Alternatively, one can also take the point $(\mathbf{u} = 5.56, \mu^* = -12)$ on the branch #3 as a remote estimate of a point on the branch #1 and determine the latter using the procedure as in section 4.1. The dotted line pt#2 (Figure 2(a)) represents the corresponding residue continuation method path to the solution on the branch #1 ($\mathbf{u} = 4.32$, $\mu = -12$). Therefore, the residue continuation scheme can also be used to switch branches. From the endpoint of pt#2, the branch #1 can again be computed with pseudo-arclength continuation. Subsequently, starting from the point ($\mathbf{u} = 2.74$, $\mu = -4$) on the branch #1, the isolated branch #2 is reached through the residue continuation path pt #3 (Figure 2(a)).

Using the residue continuation method, no specific treatment is necessary for the switch and no specific correctors are needed as is the case for predictor methods based on interpolation [1, 19, 18, 2]. Furthermore, only the operator itself and its Jacobian are needed in contrast to the predictor method via the tangent [12] where higher order derivatives are needed.

4.3. **Convergence and estimates.** In the case of a scalar operator $\mathbf{A} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, we now consider the sufficient conditions on the convergence for the residue continuation methods (9) and (10) as in the propositions 3.3 and 3.5. The local evolution of the norm of the residue versus the residue parameter is constrained by a curve similar to the function says as

$$\begin{array}{rccc} f & : & \mathbb{R} \setminus \{1\} & \to & \mathbb{R}^+ \\ & & \alpha_\nu & \mapsto & \left| \dfrac{\alpha_\nu}{1 - \alpha_\nu} \right| \end{array}, \tag{49}$$

sketched as a solid line in Figure 2(c). This curve partitions the domain $\{\alpha \in \mathbb{R} \setminus \{1\}, \mathbf{r} \geq 0\}$ in two zones. The Newton-Raphson method quadratically converges only below the curve.

For the example in section § 4.2, consider the residue path pt#3 of Figure 2(a). Along this path, the norm of the residue versus the residue parameter for the three values of the arclength step ds is plotted in

Figure 2(b). For $ds = 0.01$, the norm of the residue versus α is plotted, together with the zone of quadratic convergence, in Figure 2(c). The starting point is marked by a square ($\alpha = 1$). As already mentioned in section 2, the residue first increases ($\alpha > 1$) up to the maximum for which $\alpha = 1$. As a matter of fact, denoting $\mathbf{r}(\alpha) \equiv \mathbf{A}(\mathbf{u}(\alpha))$, it follows from the identity

$$\frac{d}{d\alpha}\|\mathbf{r}\|^2 = 2\|\mathbf{r}\|\frac{d}{d\alpha}\|\mathbf{r}\|$$

that every non zero extremum of the norm of the residue is attained for $\alpha \in I_\nu$ as defined in (27) when

$$\frac{d}{d\alpha}\|\mathbf{r}\| = 0.$$

That is, taking into account the equation (26)-(27), when either $\|\mathbf{r}_{\nu-1}\| = 0$ or $\alpha_\nu = 1$. Therefore, as long as $\mathbf{r}_\nu$ does not vanish, the extrema of the residue are located along the line $\{\alpha = 1\}$ of the graph $\|\mathbf{r}\|$ as a function of α.
Moreover: if $0 < \alpha_{\nu-1} < \alpha_\nu = 1 < \alpha_{\nu+1}$, then

$$\|\mathbf{r}_\nu\| = \|\mathbf{r}_{\nu-1}\| < \|\mathbf{r}_{\nu+1}\| \quad \text{and} \quad \|\mathbf{r}_{\nu-2}\| > \|\mathbf{r}_\nu\| = \|\mathbf{r}_{\nu-1}\|,$$

that is, if α_ν crosses $\alpha_\nu = 1$ from below, then $\|\mathbf{r}_\nu\| = \|\mathbf{r}_{\nu-1}\|$ is a local minimum. The same arguments show that, conversely, if α_ν crosses $\alpha_\nu = 1$ from above, then $\|\mathbf{r}_\nu\| = \|\mathbf{r}_{\nu-1}\|$ is a local maximum.
Generally, the sequence $(\alpha_\nu)_\nu$ is not monotonous, while $\|\mathbf{r}_\nu\|$ is monotonous as long as $|\alpha_\nu|$ remains in either the interval $]0, 1[$ or $]1, +\infty[$. For example, if $\alpha_{\nu_0} > 1$ is, at least locally, a maximum value of the parameter α_ν, then, $\|\mathbf{r}_\nu\|$ still increases beyond $\|\mathbf{r}_{\nu_0}\|$ for $\nu > \nu_0$, but at a lower rate than for $\nu < \nu_0$. Indeed, in the example above, the residue $\|\mathbf{r}_\nu\|$ (resp. α_ν) increases from the zero value (from $\alpha_\nu = 1$) until α_ν yields a local maximum $\alpha_{\max} > 1$ from which the residue still increases but at a lower rate. One observes this fold point in both the Figures 2(b) and 3(b). Then, α_ν crosses $\alpha_\nu = 1$ from above which causes the residue to decrease until $\alpha_\nu = 0$. This leads to $\mathbf{r}_\nu = 0$. Eventually, the residue decreases from the maximum down to zero. The corresponding end points of the path #3 are marked with a circle for $ds = 10^{-3}$, $ds = 10^{-2}$ and $ds = 5 \cdot 10^{-5}$ from right to left on the α- axis on the Figure 2(b). In Figures 3(a) and 3(b) the corresponding curves of the residue as a function of $\mathbf{u}$ and $\mathbf{u}$ versus α are plotted.

4.4. **Shallow truss arches.** As an illustration of the residue continuation method in a multidimensional dynamical system, we analyse the dynamics of shallow truss arches [9], as depicted in Figure 4(a). In the (x, y) plane, this two rod system is characterized by the Young

modulii E_i and the areas A_i of the sections of the rods (for $i = 1, 2$). We denote by E and A the nominal values of the perfect system. Let $(x_1 = -1, y_1 = 1)$ and $(x_2 = 1, y_2 = 1)$ be the coordinates of the stands 1 and 2. We define the lengths of the rods at the equilibrium before loading and for a given vertical load f as

$$l_i = \left((x_3 - x_i)^2 + (y_3 - y_i)^2\right)^{1/2}$$

$$\hat{l}_i = \left((x - x_i)^2 + (y - y_i)^2\right)^{1/2}$$

where (x_3, y_3) and (x, y) are the coordinates of the point 3 of the two rod system without or with a given vertical load $|f| > 0$.

The equilibrium of the arches is governed by the following system of equations

$$\begin{pmatrix} F_x \\ F_y \end{pmatrix} \equiv \begin{pmatrix} \displaystyle\sum_{i=1}^{2} \frac{E_i A_i}{EA} \left(\frac{1}{l_i} - \frac{1}{\hat{l}_i}\right)(x - x_i) & \\ \displaystyle\sum_{i=1}^{2} \frac{E_i A_i}{EA} \left(\frac{1}{l_i} - \frac{1}{\hat{l}_i}\right)(y - y_i) & - \quad f \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The vector of the unknowns is $\mathbf{u} \equiv (x, y)$ and the vector of the parameters is

$$\left(x_1, y_1, x_2, y_2, x_3, y_3, \frac{E_1 A_1}{EA}, \frac{E_2 A_2}{EA}, f\right)$$

For the perfect case, we set:

$$x_1 = -1,\ y_1 = 1,\ x_2 = 1,\ y_2 = 1,\ x_3 = 0,\ y_3 = 0,\ \frac{E_1 A_1}{EA} = \frac{E_2 A_2}{EA} = 1.$$

In this case, Figure 4(b) depicts y versus the load f. The solid and the dotted lines denote the stable and unstable parts of the branch respectively. There are two saddle-node bifurcations which lead to hysteresis behavior as a function of the load f. Starting from the point a on the branch for which $f = 0$, a vertical load ($f > 0$) yields eventually the saddle-node bifurcation at point b. For a slightly larger load, the system jumps down to the stable fixed point c. Then, if the load is decreased from point c, the system eventually reaches the saddle-node bifurcation at point d. For slightly smaller values of f, the system jumps up to the stable fixed point e. This behavior is the well known snap-through phenomenon.

For $x_3 = 0.1$, an imperfect branch #2 (Figure 4(c)) appears. Starting from a point on the perfect branch $x = 0, y = 0$ for $f = 0$ (denoted by #1 in Figure 4(c)), the residue path towards the imperfect branch

#2 is again computed using the residue continuation method. Thereafter, the imperfect branch #2 is computed using classical pseudo-arclength continuation method. In the imperfect case ($x_3 = 0.1$) hysteresis occurs both for x and y (5). In this example, the residue continuation method is an efficient tool to compute the isolated branches and hence to provide insight into the imperfect snap-through of the shallow truss arches system.

5. Summary and Discussion

Many physical systems exhibit disjointed equilibria. In this contribution, the so-called residue (homotopy) continuation method is introduced. The method is based on the Global Homotopy (6) and can start from a remote initial estimate. We have derived explicit conditions ensuring the quadratic convergence of the Newton-Raphson algorithm for the residue continuation method in both natural and pseudo-arclength continuation.

Along a few examples, the capabilities of the residue continuation method are illustrated. Isolated branches are indeed determined in quite general cases. As shown in section 4.2, the method may even be used as a branch switching algorithm. The branch switching near a transcritical bifurcation, which is classically a delicate issue, enters the scope of the method.

As an illustration of the residue continuation method in a more practical (although still low dimensional) system, we considered in section 4.4 the imperfect snap-through in a shallow two truss arch system [9]. In this example, we show that an isolated branch can be reached. The residue continuation method is likely to be a good candidate for the treatment of the imperfect bifurcations.

With respect to earlier methods suggested to determine isolated branches, a complete picture of the fixed point set is attainable with the residue continuation method without referring to any preliminary hierarchy of the singularities. Any singular points on the residue continuation path can be handled by means of a local finite step of the control or residue parameter. Finally, no higher order derivatives (only the Jacobian) are necessary for the residue continuation scheme in contrast to the predictor method via the tangent [12]. The structure of fixed point solutions is a delicate and important issue in the natural systems, in particular those that exhibit imperfections. The residue continuation is reasonably efficient, robust and easy to implement as only the operator and its Jacobian are needed.

References

[1] E.L. Allgower, K. Georg, *Numerical continuation methods*, Springer Series in Computational Mathematics, Vol. 13, 388 pp., 1990.

[2] E.L. Allgower, K. Georg, *Numerical path following*, in: P.G. Ciarlet and J.L. Lions (Eds.), Handbook of Numerical Analysis, Vol. 5 (North-Holland, Amsterdam, 1997) 818 pp., 3–208.

[3] H. A. Dijkstra, M. J. Molemaker, *Imperfections of the North-Atlantic wind-driven ocean circulation: continental geometry and assymetric wind-stress*, J. Mar. Res. 57 (1999) 1–28.

[4] C.B. Garcia, W.I. Zangwill, *Path ways to Solutions, Fixed Points, and Equilibria*, Prentice-Hall, Englewood Cliffs, NJ, 479 pp., 1981.

[5] M. Golubitsky, D. Shaeffer, *A theory for Imperfect Bifurcation via Singularity Theory*, Communications on Pure and Applied Mathematics 32 (1979), 21–98.

[6] M. Golubitsky, D. Shaeffer, *A qualitative study of the steady–state solutions for a continuous flow stirred tank chemical reactor*, SIAM. J. Math. Anal. 11 (1980) 316–339.

[7] M. Golubitsky, D. Shaeffer, *Singularities and Groups in Bifurcation Theory: I* Applied Mathematical Sciences, Vol. 51, Springer–Verlag, 463 pp., 1985.

[8] F.J. Gould, J.W. Tolle, *Complementary Pivoting on pseudomanifold Structure with applications on the Decision Sciences*, Sigma Series in Applied Mathematics, Vol. 2, Heldermann Verlag, Berlin, 202 pp., 1983.

[9] K. Ikeda, K. Murota, *Imperfect bifurcation in Structures and Materials*, Applied Mathematical Sciences, Vol. 149, Springer–Verlag, New-York, 411 pp., 2002.

[10] A. D. Jepson, A. Spence, *The numerical solution of non–linear equations having several parameters I: scalar equations*, SIAM. J. Math. Anal. 22 (1985) 736–759.

[11] L. Kantorovich, G. Akilov. Functional Analysis in Normed Spaces, Pergamo Press, Oxford, 774 pp., 1964.

[12] H. B. Keller, *Numerical solution of bifurcation and non–linear eigenvalue problem*, in: P.H. Rabinowitz (Ed.), Applications of Bifurcation Theory (Academic Press, New-York, 1977) 389 pp., 359–384.

[13] H. B. Keller, *Global homotopies and Newton methods*, in: C. deBoor, G. H. Gollub (Eds.), Recent Advances in Numerical Analysis (Academic Press, New-York, 1978) 73–94.

[14] H. B. Keller, *Lectures on Numerical Methods in Bifurcation Problems*, Springer-Verlag, Berlin, Heidelberg, New-York, 1987.

[15] Iu. A. Kuznetsov, *Elements of Applied Bifurcation Theory*, Springer Verlag, 2004.

[16] J. Ortega, W. Rheinboldt. Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New-York, 572 pp., 1970.

[17] W.C. Rheinboldt, *Numerical Analysis of Parametrized Non-Linear Equations*, J. Wiley, New-York, 299 pp., 1986.

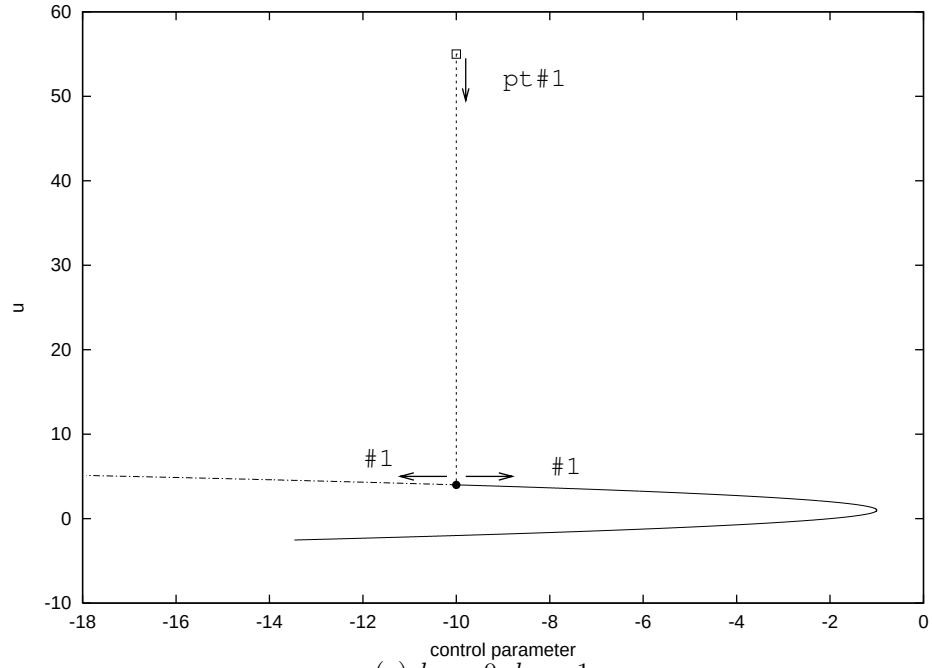
[18] R. Seydel, *Branch switching in bifurcation problems of ordinary differential equations*, Numer. Math. 41 (1983) 93-116.

[19] R. Seydel, *Practical bifurcation and stability analysis*, Springer–Verlag, 407 pp., 1994.

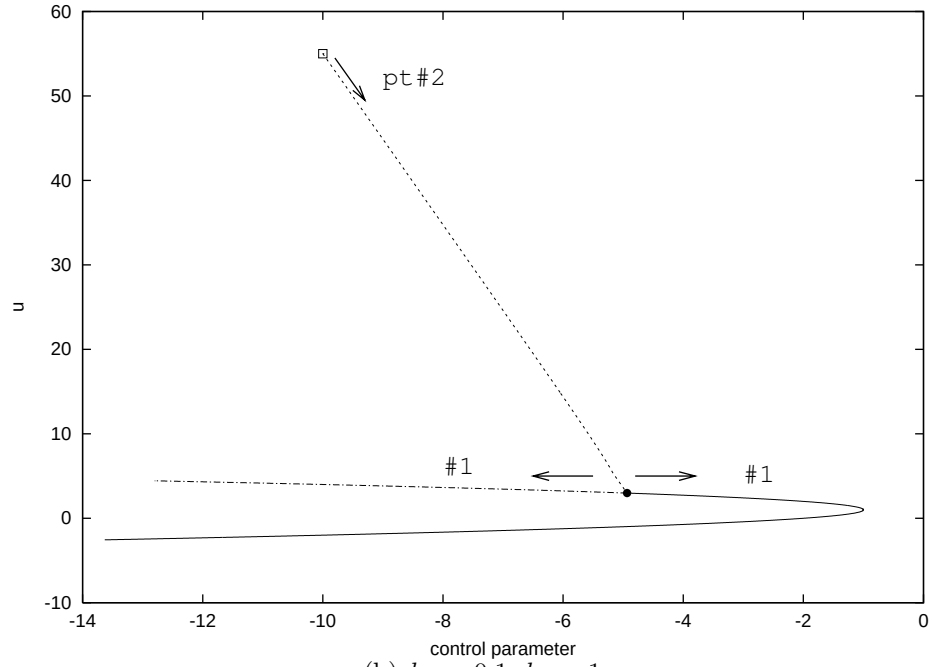
[20] E. Simonnet, M. Ghil, K. Ide, R. Temam, S. Wang, *Low-frequency variability in the Shallow-Water models of the wind-driven Ocean circulation*, J. Phys. Oceanogr. 33 (2003) 712–751.
[21] M.J. Todd, *The computation of fixed points and Applications*, Lecture Notes in Economics and Mathematical Systems, Vol. 124, Springer-Verlag, Berlin, Heidelberg, New-York, 129 pp., 1976.
[22] N. Yamaki, *Elastic stability of circular cylindrical shells*, E. Becker, B. Budiansky, W.T. Koiter and H.A. Lauwerier (Eds.), North Holland series in Applied Mathematics and Mechanics, Vol. 27, North Holland, Amsterdam, 558 pp., 1984.

Received june, 28 2005
Accepted february, 15 2006

Université de Rennes1, I.R.M.A.R, Campus de Beaulieu, 35042 Rennes Cedex (France)
E-mail address: nathalie.rittemard@univ-rennes1.fr

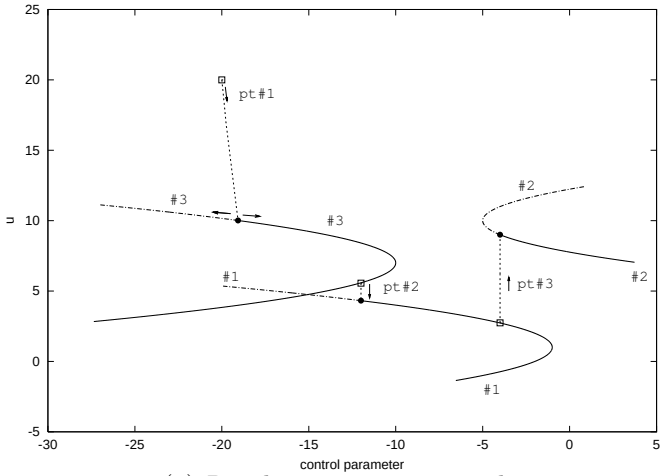


(a) $k_{\mathbf{u}} = 0$, $k_\mu = 1$

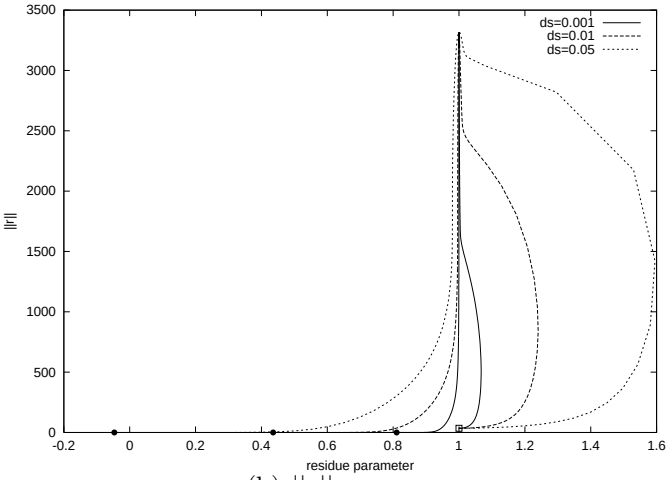


(b) $k_{\mathbf{u}} = 0.1$, $k_\mu = 1$

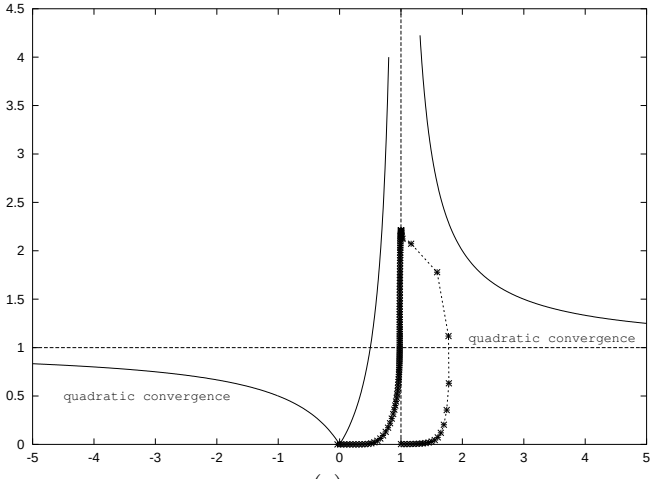
FIGURE 1. *Plot of the residue continuation path pt#1 (dotted line) from a coarse initial guess (marked with a square) of the operator (47).*



(a) Residue continuation paths

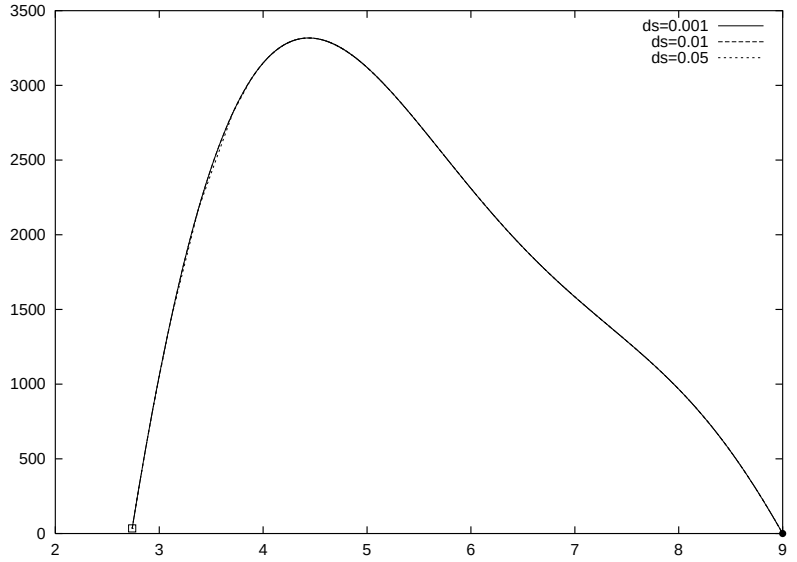


(b) $||r||$ versus α

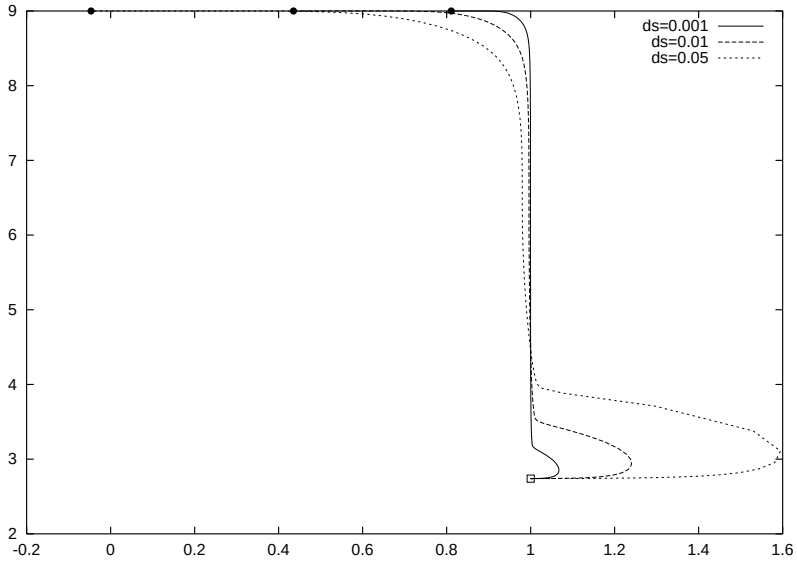


(c)

FIGURE 2. *For the operator (48), (a) continuation paths, (b) the norm of the residue versus α along pt#3 with $ds = 10^{-2}$ and (c) the sketch of the evolution of the norm of the residue (divided by 1500 for convenience) for pt#3 together with the function (49) as a solid line.*

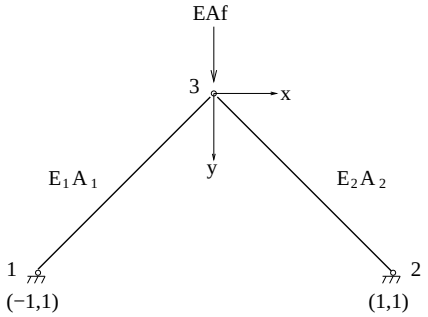


(a) $||\mathbf{r}|| = f(\mathbf{u})$

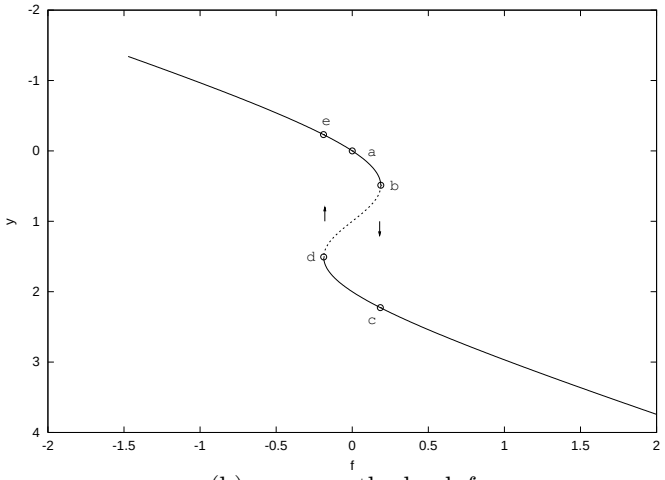


(b) $\mathbf{u} = f(\alpha)$

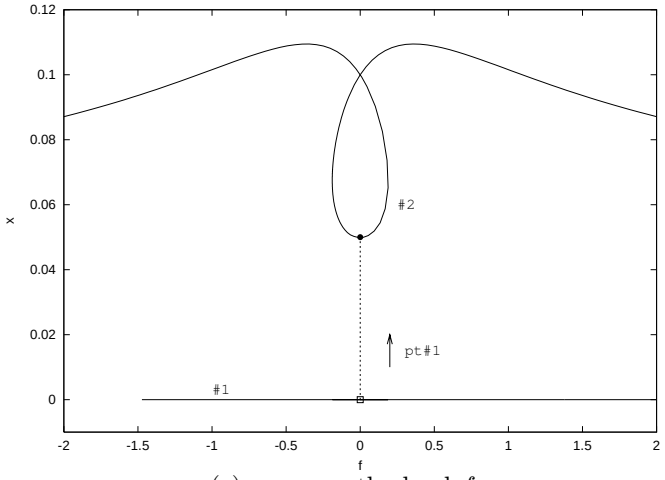
FIGURE 3. *(a) Norm of the residue versus* $\mathbf{u}$ *and (b)* $\mathbf{u}$ *as a function of the residue parameter corresponding to the Figure 2.*



(a) Residue continuation paths

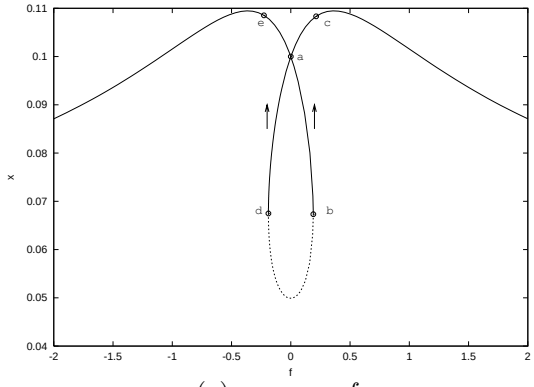


(b) y versus the load f

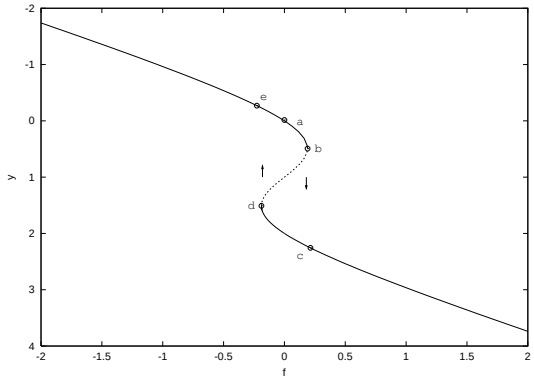


(c) x versus the load f

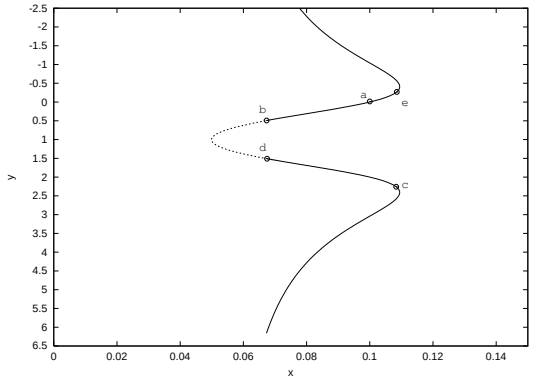
FIGURE 4. *(a) A sketch of the shallow truss arches system. (b) Bifurcation diagram for the perfect shallow truss arch system. (c) Bifurcation diagram of the imperfect ($x_3 = 0.1$) shallow truss arches system.*



(a) x versus f



(b) y versus f



(c) $y(f)$ versus $x(f)$

FIGURE 5. *Bifurcation diagram for the shallow truss arches in the imperfect case. In (a) and (b), y and x are plotted versus the load f, respectively while in (c) the parametric curve $(y(f), x(f))$ is shown. A solid and dotted linestyle denote the stable and unstable parts of the branches, respectively. In the imperfect case ($x_3 = 0.1$) hysteresis occurs both in y and x as the load f is varied.*